The many dimensions of interdisciplinarity

Pablo Jensen, Katsyarina Paradzinets

The many dimensions of laboratories' interdisciplinarity

Pablo Jensen, Katsiaryna Lutkouskaya

Institut Rhône-Alpin des Systèmes Complexes (IXXI) and Laboratoire de Physique, UMR CNRS 5672, École Normale Supérieure de Lyon, 69007 Lyon, FRANCE

Abstract : Interdisciplinarity is as trendy as it is difficult to define. Instead of trying to capture a multidimensional object with a single indicator, we propose six indicators, combining three different operationalizations of a discipline, two levels (article or laboratory) of integration of these disciplines and two measures of interdisciplinary diversity. This leads to a more meaningful characterization of the interdisciplinarity of laboratories' publication practices. Thanks to a statistical analysis of these indicators on 600 CNRS laboratories, we suggest that, besides an average value of interdisciplinarity, different laboratories can be mainly distinguished by the “distance” between the disciplines in which they publish and by the scale at which interdisciplinary integration is achieved (article or laboratory).

Introduction:

CNRS is the largest scientific organization in Europe, covering most fields of science. According to its website, “CNRS encourages collaboration between specialists from different disciplines [...] thus opening up new fields of inquiry to meet social and economic needs. CNRS has developed interdisciplinary programs which bring together various CNRS departments as well as other research institutions and industry.”

Recently, CNRS launched a “Mission for Interdisciplinarity” to “promote, facilitate and coordinate interdisciplinarity at CNRS.” This Mission aims to foster the development of new themes, or new disciplines. To this end, it commissioned a study on the quantification of interdisciplinarity of the roughly 1000 laboratories affiliated to CNRS, with the help of scientometrics. In this paper, we present the results of this study.

We start by defining a variety of indicators, to capture many facets of interdisciplinarity through publications. Interdisciplinarity means, at the most generic level, some degree of integration of different disciplines (Weingart & Stehr, 2000; Porter & Rafols, 2009; Marcovich & Shinn, 2011; Wagner et al. 2011, Rafols et al, 2012). To transform this idea into quantitative indicators, we need to answer three questions:

1. How to define a discipline?

2. At what level the integration is achieved?
3. What is the degree of disciplinary linkage achieved?

Let us briefly comment on these three points. There are several ways to define a discipline from a scientometrics' point of view. Since we are dealing with CNRS labs, the most natural would seem to use the disciplinary organization of CNRS in 10 "institutes" and 40 subdisciplinary "sections"[1]. A convenient alternative is to use the 224 Journal Subject Categories (JSCs) used by Web of Science (WoS). Finally, instead of using institutionally predefined divisions of science, one could use a more bottom-up definition of "cognitive" communities. To obtain these communities, we use the roughly 300 000 French articles published between 2007 and 2010 and group them into "cognitive communities" using clustering algorithms based on shared references. More details are given below.

We now turn to the second question, the level at which the various disciplines are connected. Again, there are several possibilities, including an article, a single author, a team and a laboratory. For example, an article may cite articles from different fields or a laboratory may gather teams or authors from different fields to work on an interdisciplinary topic. Alternatively, a single scientist may have started his career in physics and then changed to history of science or environmental sciences, or collaborate and publish with scientists from different disciplines.

Finally, the hardest question from the point of view of quantification is related to the degree of disciplinary integration achieved (Wagner et al., 2011). How to distinguish research that simply juxtaposes different disciplines, as an encyclopedia that gathers entries from different fields, from real interdisciplinary work that achieves full integration of different fields to create a new discipline? This is a key issue for understanding the cognitive aspects of interdisciplinarity (Marcovich & Shinn, 2011). Unfortunately, it is very difficult to distinguish between these different possibilities using scientometrics' indicators.

I Presentation of six indicators

I – 1 Conceptual framework

---

1 http://www.cnrs.fr/en/aboutCNRS/institutes.htm

In this paper, we will use three definitions of “discipline” and two integration levels (laboratory and article) to calculate six partial interdisciplinary indicators. Our aim is not to find the “best” indicator of interdisciplinarity but to show that interdisciplinarity has many relevant dimensions (Rafols et al, 2012; Soos & Kampis, 2012), leading to a rich description of the interdisciplinary practices of CNRS labs.

We adopt Stirling's (2007) approach to capture the different facets of diversity : ‘variety’, ‘balance’ and ‘disparity’. All else being equal, the greater each of these elements, the greater the diversity. 'Variety' characterizes the number of different categories, 'balance' characterizes the evenness of the distribution over these categories and 'disparity' characterizes the difference among the categories, usually based on some distance. For example, a laboratory that publishes articles belonging to many JSCs shows a high variety. If these articles are evenly distributed over the different JSCs, it is also characterized by a large balance indicator. Finally, its 'disparity' dimension is large if these JSCs are 'distant' in a disciplinary space (think of the difference between publishing both in Materials Science and Chemistry or in Materials Science and Psychology).

Table 1 summarizes the indicators proposed in this paper. They combine different levels of integration (laboratory, article), definitions of discipline (JSC, CNRS institute or cognitive community) and two ways of quantifying interdisciplinary diversity (Stirling, 2007). We do not show all the indicators that we have tried out, as most are strongly correlated. For example, it is possible to compute art_jsc_bal (see below, eq 3b), which turns out to be strongly correlated to lab_jsc_bal (eq 1). Therefore, as it does not bring new information, we have not included it in Table 1 or in the statistical analysis.

Disciplinary classification	***integration level***	
	article	*lab*
cognitive based : *top-down (JSC)*	art_jsc_div (4)	lab_jsc_bal (1) lab_jsc_div (2)
cognitive based : *bottom-up (articles' references)*		lab_cogn_div (7)
Organizational : CNRS institute		lab_inst_main_bal (5) lab_inst_cop_bal (6)

Table 1: Summary of the six indicators used in this article. Numbers in brackets correspond to the equation where the indicators are defined.'bal' (balance) and 'div' (diversity) refer to the way interdisciplinary diversity is quantified (Stirling 2007).

I – 2 Indicators of the diversity of JSC of a laboratory publications

These indicators are based on Journal Subject Categories (JSC) of Web of Science. From the list of a lab publications, we calculate the distribution over the JSC[2] through p_i, the proportion of articles of this laboratory in JSC_i. If a journal is related to a number n of different JSCs, it contributes a factor *1/n* to each of them. Two examples of this distribution are given in Figure *1*.

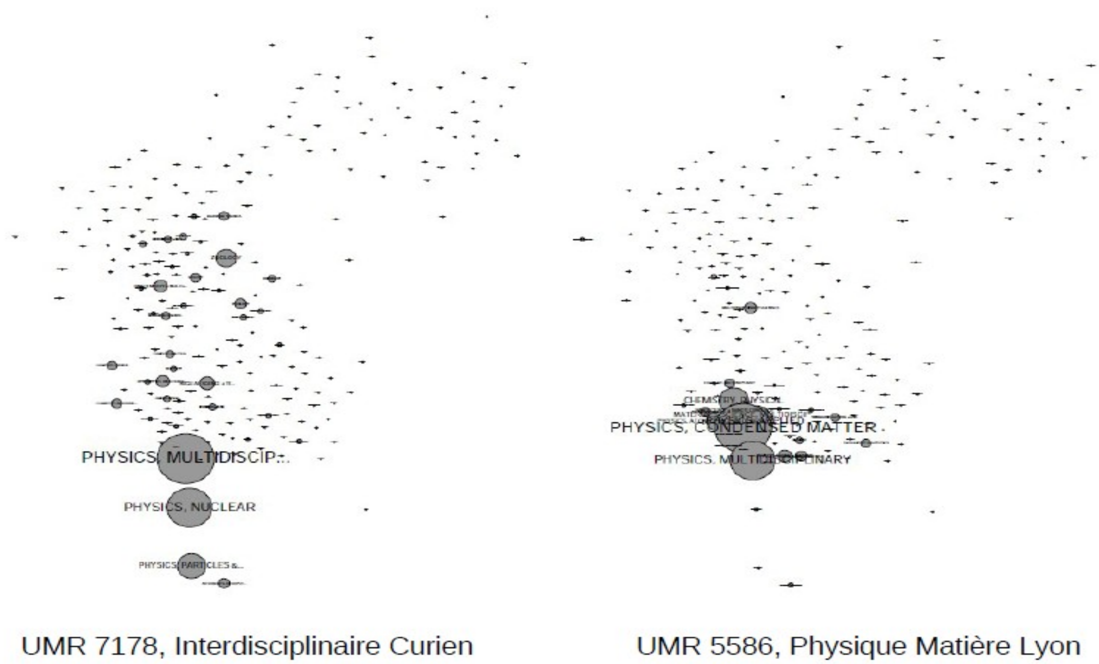

Fig. 1 *The publications of a laboratory can be represented in an "overaly map" which shows the distance between disciplines, as proposed by Leydesdorff & Rafols (2008)*

I – 2 a - Balance of the JSCs of a laboratory:

2 We exclude the JSC "Multidisciplinary Sciences" from analysis because by definition this JSC mixes articles (often monodisciplinary) from many disciplines.

A simple indicator of the spread of the disciplines where a laboratory publishes is given by:

$$lab_{jsc_{bal}} = \sum_{i,j,i \neq j} p_i p_j \qquad (1)$$

where p_i is the proportion of articles of the laboratory in JSC$_i$. This indicator does not take into account how similar or dissimilar these fields are: publishing 50% of the papers in “Applied Physics” and 50% in “Condensed Matter Physics” leads to the same value of the balance indicator as publishing 50% of the papers in ”Applied Physics” and 50% in “Sociology”.

I – 2 b - Diversity

As we would like to include the idea of “distance” between disciplines, we calculate the diversity indicator (Stirling 2007, Porter & Rafols, 2009) which combines both the spread of the disciplines through the p_i and the distance between them.

$$lab_{jsc_{div}} = \sum_{i,j,i \neq j} p_i p_j \frac{1}{s_{ij}} \qquad (2)$$

where s_{ij} is the cosine measure of similarity between JSCs *i* and *j.* Practically, s_{ij} is measured through the citations from publications in JSCs *i* to publications in JSC *j* (Porter & Rafols, 2009). When discipline *i* never quotes any article belonging to discipline *j*, s_{ij} takes the value 0, while s_{ij} is closer to 1 when such references are common. If we compare the two examples mentioned above, the first case leads to lab_jsc_div = 0.267, as the JSCs are very similar (s_{ij} = 0.935), while the second case leads to a much higher value (lab_jsc_div = 1250, since s_{ij}=0.0002 between “Applied Physics” and “Sociology”).

Note that we use 1/s_{ij} to quantify the distance between disciplines, contrary to Porter & Rafols (2009) or Roessner et al. (2013) who use (1 – s_{ij}). Clearly, using 1/s_{ij} leads to a much higher contrast between close and distant disciplines, which is what we are after here in order to detect different styles of interdisciplinarity.

I – 3 Indicators of the interdisciplinarity of single publications

A limitation of the indicators of the diversity of a lab publications is that they may show high values for a laboratory that gathers teams of different disciplines (and that, therefore, publish in different JSCs), even when there are no significant interactions between the different teams. To further characterize a lab's interdisciplinarity, it is useful to introduce an indicator of the interdisciplinarity of *single* articles, to test whether interdisciplinarity is achieved at this cognitive level.

I – 3 a Single article interdisciplinarity

The interdisciplinary diversity of a single article is calculated as that for labs (equation 2) by integrating the JSCs of the *references* listed in the article. We have obtained the JSC of a reference from the name of journal where it appears. We could identify 6,960,940 references out of 9,626,563 (after data cleaning described in http://www.sebastian-grauwin.com/*)* leading to a recovery rate of 72.3%, to be compared to the recovery rate of 80% achieved by Porter& Rafols 2009).

Specifically, the interdisciplinary diversity of a single article is calculated as:

$$art_{div} = \sum_{i,j,i\neq j} pa_i \, pa_j \frac{1}{s_{ij}} \qquad (3)$$

where pa_i is the proportion of articles' references in JSC_i. Note that we use the same definition of distance between JSC_i and JSC_j as for the laboratory level (equation 2).

In practice, we have to correct the raw values for *art_div* because WoS contains errors in the cited references. When the wrong reference belongs to a distant field (for example, to 'Sports' instead of 'Physics'), this leads to an erroneously high value of diversity. To overcome this problem, we calculate the *corrected value of diversity*:

art_div_corr = min(art_div_1) (3a)

where *art_div_1* is the "diversity without one reference". It is calculated as follows: one reference is dropped from the list of references of the article and the value of diversity is recalculated. Then, the next reference is dropped (the first reference is restored) and the second value of diversity is calculated, and so on. The minimal diversity from this set of diversities is selected as *art_div_1*. In this way, we avoid not only false references but also artificially high values of the interdisciplinarity of a single paper which would result from a single reference to a distant field. In the following, we use the corrected value for *art_div*. As the use of the distance term $1/s_{ij}$ leads to a quite volatile measure of interdisciplinarity, we have tested the consequence of removing the second most diverse reference, by a procedure similar to that followed for removing the most diverse reference. While this changes the value of the *art_div* indicator for each article, a high correlation with art_div_corr is found (0.70), which shows the statistical robustness of our indicator.

I – 3 b Balance of article:

We also compute the balance of references over the disciplinary fields, as in eq (3) by dropping the distance term $1/s_{ij}$:

$$art_{jsc_{bal}} = \sum_{i,j,i \neq j} pa_i \, pa_j \qquad (3b)$$

This indicator is expected to be more robust when looking at dynamics of interdisciplinarity over large periods of time, especially over 10 and more years, since the relationships between the disciplines, and hence the measure of similarity s_{ij}, may change in time.

I – 3 c Laboratory level indicator

To quantify the interdisciplinarity of the papers published by a lab, we aggregate the articles' diversity indicator *art_div_corr* at the laboratory level by averaging over all the articles published by that laboratory:

$$art_{jsc_{div}} = \frac{\sum_i art_{div_{corr}}}{pap} \qquad (4)$$

where *#pap* is the number of articles of the lab for which at least one reference was identified. We have also computed the aggregated value for art_jsc_bal (eq 3b). However, since it turns out to be strongly correlated to lab_jsc_bal (eq 1), we omit this indicator in the rest of the paper.

I – 4 Indicators built on CNRS' disciplinary « Institutes »

We now calculate interdisciplinarity indicators based on the 10 CNRS's disciplinary institutes: Institute of Biological Sciences (INSB), Institute of Chemistry (INC), Institute of Ecology and Environment (INEE), Institute for Humanities and Social Sciences (INSHS), Institute for Information Sciences and Technologies (INS2I), Institute for Engineering and Systems Sciences (INSIS), National Institute for Mathematical Sciences (INSMI), Institute of Physics (INP), National Institute of Nuclear and Particle Physics (IN2P3) and National Institute for Earth Sciences and Astronomy (INSU). Each laboratory is assigned to an Institute which is supposed to reflect the majority of the research fields studied by the lab. We have built two indicators based on the spread of the lab's academic production over the disciplinary Institutes.

I – 4 a- Proportion of publications outside the Institute's mainstream JSCs

First, we create a list of the most common JSCs for each institute. For this, we take all the publications of the laboratories belonging to an institute and classify them by JSC. Table 2 shows this list for Life Sciences. Then, we choose a threshold to define the most common JSCs for each institute. We have studied thresholds of 80, 90 and 99% and found that there were no significant variations in the results obtained. We therefore choose a threshold value of 90%.

	JSC	*percentage, %*	*cumulative sum, %*
1	BIOCHEMISTRY & MOLECULAR BIOLOGY	16.65	16.65
2	NEUROSCIENCES	9.13	25.78
3	CELL BIOLOGY	6.50	32.28
4	ENDOCRINOLOGY & METABOLISM	3.87	36.15
5	MICROBIOLOGY	3.83	39.97
6	GENETICS & HEREDITY	3.76	43.73
7	PLANT SCIENCES	3.34	47.07
8	MULTIDISCIPLINARY SCIENCES	3.32	50.39
9	IMMUNOLOGY	3.14	53.54
10	BIOPHYSICS	2.46	55.99
11	ONCOLOGY	2.45	58.44
12	PHARMACOLOGY & PHARMACY	2.40	60.84
13	DEVELOPMENTAL BIOLOGY	2.30	63.14
14	HEMATOLOGY	2.08	65.22
15	CLINICAL NEUROLOGY	1.99	67.20
16	VIROLOGY	1.69	68.89
17	PHYSIOLOGY	1.53	70.42
18	BIOTECHNOLOGY & APPLIED MICROBIOLOGY	1.53	71.95
19	CHEMISTRY, MULTIDISCIPLINARY	1.43	73.38
20	BIOCHEMICAL RESEARCH METHODS	1.42	74.80
21	MEDICINE, RESEARCH & EXPERIMENTAL	1.04	75.84
22	BEHAVIORAL SCIENCES	0.98	76.82
23	BIOLOGY	0.96	77.78
24	CHEMISTRY, MEDICINAL	0.89	78.66
25	CHEMISTRY, ORGANIC	0.83	79.49
26	GASTROENTEROLOGY & HEPATOLOGY	0.82	80.31
27	DERMATOLOGY	0.80	81.11
28	INFECTIOUS DISEASES	0.77	81.88
29	PSYCHIATRY	0.76	82.64
30	CHEMISTRY, PHYSICAL	0.75	83.39
31	PATHOLOGY	0.65	84.04
32	ZOOLOGY	0.59	84.63
33	REPRODUCTIVE BIOLOGY	0.58	85.21
34	CHEMISTRY, ANALYTICAL	0.57	85.78
35	TOXICOLOGY	0.56	86.34
36	RADIOLOGY, NUCLEAR MEDICINE & MEDICAL	0.55	86.88
37	IMAGING	0.54	87.42
38	PERIPHERAL VASCULAR DISEASE	0.51	87.93

39	RHEUMATOLOGY	0.40	88.33
40	CARDIAC & CARDIOVASCULAR SYSTEMS	0.40	88.74
41	FOOD SCIENCE & TECHNOLOGY	0.39	89.12
42	PARASITOLOGY	0.37	89.49
43	MEDICINE, GENERAL & INTERNAL	0.36	89.86
	VETERINARY SCIENCES		

***Table 2**: List of the most common JSCs for CNRS's Biology Institute (INSB) for the threshold of 90%*

Then, for each laboratory, we count the percentage of articles *outside* this 90% list and normalize by the expected value, i.e. the average value 0.1.

$\bar{s}_{INST}90$

$i,JSC \notin JSC$

$$\sum_{i}^{p} \qquad (5)$$

$lab_{inst_{main_{bal}}}$

where $p_{i,\ JSC \notin JSCs\_INST90}$ are the frequencies of the JSCs that do not belong to the Institute's JSC main list. Since this indicator relies only on the proportions of JSCs, without a notion of distance, it is akin to a 'balance' indicator of interdisciplinarity.

I – 4 b- Proportion of co-publications with CNRS laboratories belonging to different Institutes

Interdisciplinary collaborations can also be detected by copublications between scientists belonging to different CNRS Institutes. We compute a fifth indicator by calculating the proportion of a lab's publications that involve authors from other Institutes :

$$lab_{inst_{cop_{bal}}} = \frac{\sum_i article_{i,nb_{INST}>1}}{articles} \qquad (6)$$

where the sum counts the number of articles of the lab involving more than one institute and *#articles* is the total number of articles published by the laboratory.

I – 5 Indicator built on cognitive disciplines

Finally, we use an alternative definition of discipline, not institutional as before (WoS or CNRS), but emerging from the articles themselves. To build these “cognitive disciplines”, we use bibliographic coupling (Kessler, 1963) between the 300 000 papers published by French laboratories in the period 2007-2010 and compiled by the Web of Science. Links between pairs of articles are calculated through their common references (at least two references are needed to create a link, to avoid non-significant links based on general references). Links are weighted, i.e. we define a similarity between two articles *i* and *j* as the cosine distance:

$$w_{ij}=\frac{\overset{refs_i}{common_{refs}}}{\sqrt{refs_j}}$$

where $\#common\_refs_{ij}$ is the number of common references for articles *i* and *j*, and $\#refs_i$, $\#refs_j$ are the numbers of references of articles *i* and *j*, respectively. By definition, $w_{ij} \in [0,1]$ is equal to zero when *i* and *j* do not share any reference, and to 1 when their sets of references are identical. In comparison to a co-citation link (which is the usual measure of articles’ similarity), bibliographic coupling (BC) offers two advantages: it allows to map recent papers (which have not yet been cited) and it deals with all published papers (whether cited or not). The reason why weighted links are used is that they reinforce the dense (in terms of links per article) regions of the BC networks. This reinforcement facilitates the partition of the network into meaningful groups of cohesive articles, or communities. A widely used criterion to measure the quality of a partition is the modularity function (Fortunato & Barthélémy 2007), which is roughly the number of edges 'inside communities' (as opposed to 'between communities'), minus the expected number of such edges if the partition were randomly produced. We compute the graph partition using the efficient heuristic algorithm presented in (Blondel et al. 2008). The whole method is described in (Grauwin & Jensen, 2011).

Applying this algorithm yields a partition of French papers into roughly 250 communities containing more than 100 papers each. Simple frequency analysis then allows to characterize each community through its more frequent items (keywords, authors, etc...). Once these cognitive disciplines are created, for each lab we compute the proportion of its publications in community i (*p_i)*. The distance between communities is set as the inverse of the mean bibliographic coupling weight between articles in communities *I* and *J*:

$$\langle w \rangle_{IJ}^{-1} = \langle w_{ij} \rangle_{i \in I, j \in J}^{-1} = \left(\Omega_{IJ} / N_I N_J \right)^{-1}$$

where Ω_{ij} is the is the total weight of the links between communities I and J, N_I and N_J the number of articles in communities I and J respectively. By analogy with the diversity formula based on JSC (formula 2), we define the diversity across "cognitive communities" as:

$$lab_{cogn_{div}} = \sum_{i,j,i \neq j} p_i p_j \frac{1}{s_{ij}} \qquad (7)$$

where *p_i* is and *p_j* are the proportions of the labs' papers belonging to communities *i* and *j* respectively.

For the sake of transparency, and to allow other researchers to use it for further analyses, the table containing the six indicators for the 671 laboratories is available as supplementary information.

II Results

II 1 Orders of magnitude: how interdisciplinary are average papers?

Before analyzing in detail the results obtained for the laboratories, it is useful to compute several orders of magnitudes on the range of disciplines used, on average, by papers and laboratories.

On average, articles refer to papers from almost 10 different disciplines (9.8 JSC). We only include in this calculation articles citing more than 10 references, which represent 2/3 of the French articles, with an average number of references equal to 32.5. However, when considering those JSC that are used in more than 10% of the reference list, this average drops to 2.7. This means that, on average, an article spreads its references on 3 main JSCs and 7 additional which benefit from roughly a single reference.

An average laboratory publishes in journals belonging to 34 different JSCs, a figure which shrinks significantly to 2.5 distinct JSC when imposing the same threshold of 10%. These figures are calculated by count only the 680 laboratories that have published more than 50 papers over the period 2007-10.

Finally, it is interesting to note that, in agreement to other quantitative studies (Porter & Rafols, 2009; Larivière & Gingras 2013), there seems to be only a small increase in the interdisciplinarity of articles, as shown in Figure 2. This slow increase may reflect a progressive drift of scientific practices away from fixed JSC, without any real crossing of cognitive frontiers.

Fig. 2 Evolution of the interdisciplinary diversity of articles over the last 20 years. We plot both the diversity including disciplinary distances (art_div_corr) and not including them (art_jsc_bal) (equations 3a & 3b). For the sake of visualization, we normalize the indicator values to 1 for 1990. Both indicators of interdisciplinary suggest a small increase of papers' interdisciplinarity over the years, as the high 1990 value for art_div_corr is not significantly higher than the 2000 value, as shown by the error bars which refer to 1.96 times the standard deviation of the mean (95% significance test).

II 2 Laboratory indicators

We have computed the six indicators for the 680 laboratories which have published more than 50 papers over 2007-2010. To allow comparisons and statistical analysis, since the absolute values have no intrinsic meaning, we have scaled all the values to achieve an average value of 0 and a variance of 1. We then carried out a Principal Component Analysis of the (680 × 6) matrix using the free software R (www.r-project.org/). More precisely, we used *prcomp* from the 'stats' package, without any axes rotation. The indicators used for the analysis are summarized in the Table 1.

II 2 a PCA analysis

Table 3 shows the coordinates of the first four axes of the PCA. The respective cumulative variances are : 0.37; 0.56; 0.71 and 0.83.

Indicator	*PC1*	*PC2*	*PC3*	*PC4*
lab_jsc_bal	0.34	-0.53	0.29	-0.65
lab_inst_cop_bal	0.38	-0.38	0.36	0.74
lab_inst_main_bal	0.49	-0.20	-0.24	-0.04
lab_cogn_div	0.33	0.55	0.54	-0.15
art_jsc_div	0.42	0.06	-0.65	0.01
lab_jsc_div	0.47	0.47	-0.05	0.01

Table 3 Coordinates of the first four axes of the PCA analysis

Let us now discuss the principal dimensions that emerge from this analysis, together with some representative labs.

II 2 b Some representative examples

II 2 b 1 PCA1: Combined interdisciplinarity

The main axis represents a combination of the various interdisciplinarity indicators. A representative example of a laboratory with low interdisciplinarity is UMR7652, an Organic Synthesis laboratory from the Chemistry Institute located at Ecole Polytechnique (pca1=-3.1). All its six interdisciplinarity indicators are well below average values. The lab publications are highly centered on the Chemistry JSCs, with only one publication out of 79 published in a journal not entirely related to Chemistry (JSC = ”Pharmacology”). Similarly, 50 of the 51 publications of the “Observatoire Aquitain” (UMS2567) are within a single JSC ”Astronomy and Astrophysics” (pca1=-3.5).

On the other side of the spectrum, Laboratoire Interdisciplinaire Hubert Curien (UMR7178) has the strongest pca1 (+7.7). It has also the strongest lab_cogn_div and lab_jsc_div because it publishes papers on very distant JSCs: 143 papers on “Physics, Nuclear”, 50 on “Zoology”, 27 on “Chemistry, Multidisciplinary”... over a total of 929 papers. However, its papers are less interdisciplinary than average (art_jsc_div = -0.25), pointing to a low degree of integration of its different teams from different disciplines, as will be detected by PCA3 (see below). Another very interdisciplinary lab is TIMC-IMAG (Techniques for biomedical engineering and complexity management, UMR 5525, pca1=+6.8). Out of its 304 papers, 24 were in published in JSC=”Neurosciences”, 8 in ”Genetics & Heredity”, 7 in “Sports Sciences” and 4 in “Mathematics”.

II 2 b 2 PCA2: Short or long cognitive distance

This axis distinguishes those labs that connect distant or nearby disciplines. For example, the “Center for the study of divided matter”, UMR 6619, is quite interdisciplinary (pca1=+1.5) but on rather short distances (pca2=-1.7), as its main JSC indicate: “Chemistry, Physical” (19%), “Materials Science” (14%), “Physics, Atomic” (7%). On the contrary, the “Center of research on the mathematics of decision”, UMR 7534, with a similar overall interdisciplinarity (pca1=+2.2) reaches longer distances (pca2= +3.3), as its JSC include “Astrophysics” and the very distant “Biology”, or “Automation”. Note that the calculation of the diversity including as “distance” the inverse cosine measure ($1/s_{ij}$) strongly increases the contribution of distant fields, even if their representation in the publications' list of a lab is small. In this last case, "Astrophysics", "Biology" or "Automation" represent each 1% of the lab's publications, but contribute hugely to its total integration, as they are very distant (huge $1/s_{ij}$).

II 2 b 3 PCA3: article or laboratory interdisciplinarity

This axis distinguishes labs that achieve interdisciplinarity either at the laboratory or article level. The first category publishes rather mono-disciplinary articles (low art_jsc_div) but in journals from different (and distant) JSCs (high lab_jsc_div). The opposite category publishes interdisciplinary articles (high art_jsc_div) in a few (and related) JSCs (low lab_jsc_div). The two extremes are given by Laboratoire Interdisciplinaire Hubert Curien (UMR7178). As we have seen previously, it has the strongest pca1 (+7.7) but also the strongest pca3 (+4.5), arising from its low art_jsc_div (-0.25). This means that its papers are mostly mono-disciplinary, pointing to a lab constituted of a juxtaposition of rather mono-disciplinary teams, each publishing in its own field. For this kind of laboratories, the average interdisciplinarity suggested by PCA1 may well be misleading, as all our indicators at the laboratory consider these teams as a single entity (the laboratory), as if they had managed to integrate knowledge across the disciplinary teams, an unwarranted assumption... On the opposite side, the lab "Natural history of the prehistoric man", UMR 7194, is also interdisciplinary on average (pca1=+3.7) but has a strongly negative pca3 (-6.0), arising from a small lab_cogn_div (-0.51) and a huge art_jsc (+9.3). UMR 7194 publishes in journals linked to its disciplinary heart ("Evolutionary Biology"). However, its papers gather references not only to this field but also to the analytical tools used to analyze objects such as teeth, the tools coming from other disciplines such as "Chemistry" or "Physics" which are quite distant from "Evolutionary Biology" or "Paleontology". This gives rise to highly interdisciplinary papers as quantified by art_jsc_div.

II 2 b 4 PCA4: Diversity of publications' JSCs or diversity of collaborations

This axis distinguishes labs that publish in journals belonging to different JSCs (high lab_jsc_bal) from labs that co-publish with labs from different CNRS Institutes (high lab_inst_cop_bal). Consider, for example, UMR7194 ("Natural history of the prehistory man") and UMR8151 ("Genetic and Chemical Pharmacology and Imagery"): they show approximately the same average interdisciplinarity (PCA1 equal to +3.7 and +1.7, respectively) but they appear as opposed along the PCA4 axis. The interpretation is that the first lab, as we have seen (II 2 b 3), publishes in a few JSCs (low lab_jsc_bal) but co-publishes with many labs from other Institutes, while the Genetics lab does the reverse: it publishes in many JSCs (more than 2 papers on 19 different JSCs) but collaborates with only a few labs outside its own Institute.

III Discussion & Perspectives

What have we learnt about the interdisciplinarity of CNRS labs? First, let us note that using the first four PCA axes gives an overall view about the interdisciplinarity practices of each lab. This view has been compared to expert knowledge, namely scientists working in those labs or scientific advisors from CNRS. This comparison, carried out for about 20 different labs from all the disciplines, suggests that these indicators characterize interdisciplinarity in a meaningful way. For example, CNRS advisors recognized that the interdisciplinarity claimed by the "Laboratoire Interdisciplinaire Hubert Curien (UMR7178)" was ongoing work, as teams from the different disciplines had not (yet) achieved knowledge integration, confirming the diagnosis based on the indicators. They also confirmed that the Organic Synthesis laboratory from the Chemistry Institute located at Ecole Polytechnique (UMR7652) is a very monodisciplinary (yet scientifically outstanding) laboratory. Clearly, more rigorous tests are needed to validate our work.

A major drawback of our method is that we cannot distinguish real *inte*rdisciplinary collaborations, giving rise to new concepts or to a coherent new scientific field, from simple *pluri*disciplinary practices that merely juxtapose different disciplines, as when historians use characterizing tools from physics. It seems difficult to learn much about the cognitive dimensions of interdisciplinarity from an automatic analysis of metadata of the papers.

There are open questions that could be addressed by further studies:

- it would be interesting to use the 'intermediation' approach to interdisciplinarity, as suggested by Leydesdorff (2007). We could for example create a network of copublications between laboratories and then quantify the betweenness centrality of each lab to detect those that seem to play the role of intermediaries.
- what can actors (in this case, scientists) from the field learn from this analysis? It is reasonable to assume that they are already aware of the degree of interdisciplinarity of their own labs. Possibly, they can learn from the comparison with the results obtained by other labs, from the same discipline or not.
- How do CNRS policy officers use these data or what do they learn? How do they update their policy?

Acknowledgements: We thank the “Service d'appui à la politique et à la prospective scientifiques” of the CNRS provided publication data related to CNRS laboratories for our research work. We also thank the CNRS Interdisciplinary Mission for financial support.

References:

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, P10008.

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences USA*, *104*, 36–41.

Grauwin, S., & Jensen P. (2011). Mapping scientific institutions. *Scientometric*s, *89*, 943-954.

Kessler, M. M. (1963). Bibliographic coupling between scientific articles. *American Documentation, 24*, 123–131.

Larivière, V., & Gingras, Y. (2013). Measuring Interdisciplinarity over the Course of a Century. *to be published in B. Cronin & C. Sugimoto (Eds.), Next Generation Metrics: Harnessing Multidimensional Indicators of Scholarly Performance. Cambridge MA: MIT Press.*

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Soceity for Information Science and Technology*, *58*(9), 1303–1319.

Leydesdorff, L., & Rafols, I. (2008). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*.

Marcovich, A., & Shinn, T. (2011). Where is disciplinarity going? Meeting on the borderland. *Social Science Information, 50*(3–4), 582–606.

Porter, A. L. & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81,* 719–745.

Rafols, I., Leydesdorff, L., O’Hare, A., Nightingale, P., & Stirling, A. (2011). How journal rankings can suppress interdisciplinarity. The case of innovation studies in business and management. *Research Policy*, *12*, 61–90.

Roessner, D., Porter, A. L., Nersessian, N. J. & Carley, S. (2013). Validating indicators of interdisciplinarity: linking bibliometric measures to studies of engineering research labs. *Scientometrics, 94*, 439–468.

Soós, S., Kampis, G., 2012. Beyond the basemap of science: mapping multiple structures in research portfolios: evidence from Hungary. *Scientometrics*, *93*, 869–891.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, *4*(15), 707–719.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature. *Journal of Informetrics*, *5*(1), 14–26.

Weingart, P., & Stehr, N. (2000). Practicing interdisciplinarity. *Toronto: University of Toronto Press*.